

# The Impact of Three Evidence-Based Programmes Delivered in Public Systems in Birmingham, UK

Michael Little, The Social Research Unit, Dartington, United Kingdom  
 Vashti Berry, The Social Research Unit, Dartington, United Kingdom  
 Louise Morpeth, The Social Research Unit, Dartington, United Kingdom  
 Sarah Blower, The Social Research Unit, Dartington, United Kingdom  
 Nick Axford, The Social Research Unit, Dartington, United Kingdom  
 Rod Taylor, University of Exeter Medical School, United Kingdom  
 Tracey Bywater, Institute for Effective Education, University of York, United Kingdom  
 Minna Lehtonen, The Social Research Unit, Dartington, United Kingdom  
 Kate Tobin, The Social Research Unit, Dartington, United Kingdom

## Vol. 6 (2) 2012

### Focus:

#### Evidence-based Developmental Prevention of Bullying and Violence in Europe

#### Editorial (p. 165)

#### Guest Editorial: The Future of Research on Evidence-based Developmental Violence Prevention in Europe –

#### Introduction to the Focus Section Manuel Eisner / Tina Malti (pp. 166 – 175)

#### Violence Prevention in Austrian Schools: Implementation and Evaluation of a National Strategy Christiane Spiel / Petra Wagner / Dagmar Strohmeier (pp. 176 – 186)

#### Clinical Significance of Parent Training for Children with Conduct Problems Martin Forster / Åsa Kling / Knut Sundell (pp. 187 – 200)

#### From Clinical-Developmental Theory to Assessment: The Holistic Student Assessment Tool Gil Noam / Tina Malti / Martin Guhn (pp. 201 – 213)

#### Preventing Child Behavior Problems at Preschool Age: The Erlangen-Nuremberg Development and Prevention Study Friedrich Lösel / Mark Stemmler (pp. 214 – 224)

#### Introducing, Researching, and Disseminating the Incredible Years Programmes in Wales Judy Hutchings (pp. 225 – 233)

#### Implementation of PATHS Through Dutch Municipal Health Services: A Quasi-Experiment Ferry X. Goossens / Evelien M. J. C. Gooren / Bram Orobio de Castro / Kees W. van Overveld / Goof J. Buijs / Karin Monshouwer / Simone A. Onrust / Theo G. W. M. Paulussen (pp. 234 – 248)

#### Effectiveness of a Universal School-Based Social Competence Program: The Role of Child Characteristics and Economic Factors Tina Malti / Denis Ribeaud / Manuel Eisner (pp. 249 – 259)

#### ► The Impact of Three Evidence-Based Programmes Delivered in Public Systems in Birmingham, UK Michael Little / Vashti Berry / Louise Morpeth / Sarah Blower / Nick Axford / Rod Taylor / Tracey Bywater / Minna Lehtonen / Kate Tobin (pp. 260 – 272)

#### Successful Bullying Prevention Programs: Influence of Research Design, Implementation Features, and Program Components Bryanna Hahn Fox / David P. Farrington / Maria M. Ttofi (pp. 273 – 282)

#### Tackling Cyberbullying: Review of Empirical Evidence Regarding Successful Responses by Students, Parents, and Schools Sonja Perren / Lucie Corcoran / Helen Cowie / Francine Dehue / D'Jamila Garcia / Conor Mc Guckin / Anna Sevcikova / Panayiota Tsatsou / Trijntje Völlink (pp. 283 – 292)

#### KiVa Antibullying Program: Overview of Evaluation Studies Based on a Randomized Controlled Trial and National Rollout in Finland Christina Salmivalli / Elisa Poskiparta (pp. 293 – 301)

#### Knowing, Building and Living Together on Internet and Social Networks: The ConRed Cyberbullying Prevention Program Rosario Ortega-Ruiz / Rosario Del Rey / José A. Casas (pp. 302 – 312)

#### Empowering Students Against Bullying and Cyberbullying: Evaluation of an Italian Peer-led Model Ersilia Menesini / Annalaura Nocentini / Benedetta Emanuela Palladino (pp. 313 – 320)

#### Identity Centrality and In-Group Superiority Differentially Predict Reactions to Historical Victimization and Harm Doing Rezarta Bilali (pp. 321 – 337)

#### A Farewell to Innocence? African Youth and Violence in the Twenty-First Century Charles Ugochukwu Ukeje / Akin Iwilde (pp. 338 – 350)

### Open Section



This work is licensed under the Creative Commons Attribution-NoDerivatives License.

ISSN: 1864-1385

# The Impact of Three Evidence-Based Programmes Delivered in Public Systems in Birmingham, UK

Michael Little, The Social Research Unit, Dartington, United Kingdom  
Vashti Berry, The Social Research Unit, Dartington, United Kingdom  
Louise Morpeth, The Social Research Unit, Dartington, United Kingdom  
Sarah Blower, The Social Research Unit, Dartington, United Kingdom  
Nick Axford, The Social Research Unit, Dartington, United Kingdom  
Rod Taylor, University of Exeter Medical School, United Kingdom  
Tracey Bywater, Institute for Effective Education, University of York, United Kingdom  
Minna Lehtonen, The Social Research Unit, Dartington, United Kingdom  
Kate Tobin, The Social Research Unit, Dartington, United Kingdom

The Birmingham Brighter Futures strategy was informed by epidemiological data on child well-being and evidence on “what works,” and included the implementation and evaluation of three evidence-based programmes in regular children’s services systems, as well as an integrated prospective cost-effectiveness analysis (reported elsewhere). A randomised controlled trial (RCT) of the Incredible Years BASIC parenting programme involved 161 children aged three and four at risk of a social-emotional or behavioural disorder. An RCT of the universal PATHS social-emotional learning curriculum involved children aged four–six years in 56 primary schools. An RCT of the Level 4 Group Triple-P parenting programme involved parents of 146 children aged four–nine years with potential social-emotional or behavioural disorders. All three studies used validated standardised measures. Both parenting programme trials used parent-completed measures of child and parenting behaviour. The school-based trial used teacher reports of children’s behaviour, emotions, and social competence. Incredible Years yielded reductions in negative parenting behaviours among parents, reductions in child behaviour problems, and improvements in children’s relationships. In the PATHS trial, modest improvements in emotional health and behavioural development after one year disappeared by the end of year two. There were no effects for Triple-P. Much can be learned from the strengths and limitations of the Birmingham experience.

There is good reason to be concerned about the well-being of children in the United Kingdom. Research using repeated application of the same measures charted a downward trend in mental health outcomes over a three-decade period (Collishaw et al. 2004), and while conduct problems reached a plateau in the early 2000s (albeit at a level that is still concerning), this pattern is not so clear for emotional problems (Maughan et al. 2008). Indeed, the

proportion of young people reporting frequent feelings of depression or anxiety doubled between 1986 and 2006 (Collishaw et al. 2010). Cross-national comparisons using a range of indicators, including material well-being, education, behaviour, and family and peer relationships, find that children in the United Kingdom regularly perform poorly compared with other high-income nations (e.g. UNICEF 2007).

The authors would like to acknowledge the team of people at Birmingham City Council involved in this project, who have demonstrated that public services can be forward-thinking. Particular note should be given to Stephen Hughes and Cheryl Hopkins for their leadership. We also acknowledge the input from the Centre for Health Economics and Medi-

cines Evaluation at Bangor University, in particular Rhiannon Tudor Edwards and Pat Linck. All three trials were registered with the Current Controlled Trials Registry (ISRCTN): Incredible Years (ISRCTN 48762440); Triple-P (ISRCTN 10429692); and PATHS (ISRCTN32534848). The randomisation sequence for all three trials were cre-

ated by the North Wales Organisation for Randomised Trials in Health (& Social Care) (NWORTH). Interests declared: none. Ethical clearance: The Warren House Group Ethics Committee and the North Wales NREC (reference 10/WNo01/29).

A number of UK government initiatives from 1998 onwards sought to reverse this trend, including an assault on child poverty, the introduction of universal pre-school provision (Eisenstadt 2011), a drive to improve school performance, and efforts to better integrate social care, youth justice, education and mental health services. For example, in 2004, local authority services for children were integrated, creating single “children’s services” departments with one director and an elected politician who is accountable for children’s health and development.

There are 150 local authorities in England. The work described in this article took place in the largest, Birmingham, serving a child population of 260,000 with a budget of roughly £1.3 billion (€1.6 billion). Birmingham is one of a small number of English local authorities where the majority of children are from minority ethnic groups. The city excels in many areas, for example school performance, but there are significant risks to well-being, not least from poverty.

### 1. Bridging Science and Policy

The Social Research Unit at Dartington is an independent centre dedicated to improving child outcomes through research, service development, and dissemination. The methods and tools it uses are underpinned by an approach known as *Common Language*. This provides a conceptual framework or way of thinking that can allow people with different roles – say policy makers and scientists – or different disciplines – such as education, psychology, social work – to communicate effectively about how to improve the well-being of children. The approach includes various elements, including epidemiology, strategy development, service design, and training (see Axford and Morpeth 2012).

In Birmingham, *Common Language* was used to bring together leaders of children’s services to produce a single vision for children, clearly specifying the outcomes they wanted to achieve. The work required that the system leaders identify those activities most likely to deliver target outcomes, and that they be clear about the source of investments necessary to fund the activities. The *Common Language* methods ensured that this emerging strategy was backed up by strong logic; the best evidence on the well-

being of local children, obtained through an epidemiological study conducted in the city as part of the Common Language work (Axford and Hobbs 2011; Hobbs et al. 2011; Axford et al. 2012); and reliable information on what works, for whom, when, and why.

These methods have been developed over a decade and used to support both the integration of children’s services in Norfolk in the East of England and, later, a \$200 million (€156 million) philanthropic investment to improve outcomes for children in Ireland (Little and Abunimah 2007; Axford et al. 2008). The tools have been refined with each application.

### 2. Brighter Futures Strategy

The move to create single departments of children’s services led Birmingham to appoint a new strategic director, Tony Howell. He decided to take an inclusive approach to strategy development by involving all of the agencies working with children, including the third sector. The Social Research Unit at Dartington was commissioned to facilitate the strategy development using *Common Language*.

The result was a single strategy for the City known as “Brighter Futures” (BCC 2007). It prioritised six outcomes, including behaviour and emotional well-being. Brighter Futures supported increased use of evidence-based programmes with proven impact on child outcomes, and stressed the need to improve parenting.

Financial support for the strategy came from the “business transformation” applied by the City Council to public services across the city, which sought to encourage investments that would later generate savings in expenditure, for example building new low-maintenance, energy-efficient offices to replace older and more expensive ones. In the context of Brighter Futures, data from a range of international sources, strongly influenced by the work of the Washington State Institute for Public Policy (Aos et al. 2004), was used to calculate that a £42 million (€52 million) investment that included the evidence-based programmes as well as substantial expenditure on staff development and IT systems would produce an economic return of £101 million (€126 million) over a fifteen-year period.

The strategy development and the requirement to produce a return on investment created strong support for evidence-based programmes. The Social Research Unit used its Common Language approach to facilitate teams of systems leaders to select a portfolio of programmes that crossed several stages of child development and had the greatest impact on outcomes targeted by Brighter Futures. Four were eventually selected.

Family Nurse Partnership (Nurse Family Partnership in the US) was offered to high-risk and usually teenage mothers beginning pre-birth (Olds 1998/2006). Success is measured in terms of improved parenting, better child health and development, delay in the birth of a second child, and improved family income.

The Incredible Years BASIC parenting programme (Webster-Stratton 1994) was offered to the parents of three- and four-year-olds showing the symptoms of a conduct disorder (e.g. often having temper tantrums, often fighting with other children, being spiteful to other children, often argumentative with adults). Building on the work of Judy Hutchings in Wales (Hutchings et al. 2007; Hutchings, Bywater, and Daley 2007), the twelve-week version of this evidence-based programme was delivered in children's centres (the universal pre-school provision that became available in the United Kingdom from 2000 onwards). The success of Incredible Years is measured in terms of improved behaviour and better social relationships at home and with other children. The programme was delivered by a mixture of children's centre staff, family support workers, educational psychologists, and parenting practitioners. All implementation staff were trained by an accredited Incredible Years trainer (Judy Hutchings) and participated in weekly half-day supervision sessions with that trainer.

Promoting Alternative Thinking Strategies (PATHS) is a school-wide curriculum to improve regulation of emotions for children aged four–eleven (Greenberg and Kusché 2002), which in Birmingham took up one hour per week for three of the primary school years (reception, year one and year two). Success is measured in terms of better behaviour and improved emotional well-being. Systematic reviews, for example by Durlak and colleagues (2011), sug-

gest that social and emotional learning programmes like PATHS produce better-behaved, happier children who therefore learn more, generating increases in academic attainment. Teachers implementing PATHS received two days of training (initial and top-up) and technical assistance from trained PATHS coaches.

Triple-P, like Incredible Years, is a parenting programme. In Birmingham the Standard Level-4 Group Triple-P parenting programme (Bodenman et al. 2008) was offered to families with children aged four–nine years whose significantly poor behaviour was causing problems at home, in school and, for some, in the community. Unlike the other programmes, the implementation team inherited practitioners with varying degrees of training in Triple-P. These were a mixture of clinical psychologists and mental health practitioners as well as parenting coordinators. These practitioners received “top-up” support from the programme originator Matt Sanders. Outcomes for Triple-P are measured in terms of children's behaviour, aggression, and emotional well-being. Supervision is not mandatory for Triple-P facilitators, although online support is offered by Triple-P International Ltd. and Birmingham facilitators also had access to additional support from a nominated Triple-P trainer within the Parenting Support Service, if they felt they required it.

The Brighter Futures strategy was rooted in high-quality epidemiological data about the well-being of children living in the city (Hobbs et al. 2011). This data was gathered on over 5,000 children through a representative school-based survey of seven–eighteen year-olds and a representative household survey of parents of zero–six year-olds, and identified aspects of children's well-being in need of particular attention, for example because on standardised measures of health and development children were performing below national norms. The strategy also drew heavily on international data on “what works”, particularly as captured in online clearing-houses of effective programmes, such as Blueprints for Violence Prevention (<http://www.colorado.edu/cspv/blueprints/>). The business transformation method required proof of return on investment. These forces led Birmingham to radically change its approach to evaluation. It subjected all four of the evidence-based programmes to

Randomised Controlled Trials (RCTs), summarised later in this article, a major departure from traditional local authority approaches to research and development.

All of the evaluations applied the “intention to treat” principle, meaning that results include those children, parents, or schools that dropped out of the study. The findings therefore reflect what happens in real-world situations, with many intervention recipients either not starting or not completing an intervention paid for by the local authority. Each of the trials used a “waiting list” design, meaning that children or schools not receiving the intervention were given priority to receive it in future if the results of the evaluation were positive. Children in the control conditions received “services as usual”, which in some cases involved substantial support – for example, the SEAL (Social and Emotional Aspects of Learning) programme in the case of the PATHS trial. Participants in the programme groups could also continue to receive services as usual – that is, no services were withdrawn – although it is acknowledged that logistically this may have been difficult (for example, if PATHS lessons used curriculum time previously allocated to SEAL).

Typically, experimental evaluation is expensive. In order to reduce costs, the Social Research Unit sought only to replicate the findings established in other trials, thereby collecting considerably less data than is usually the case. The experimental approach was taken, randomly allocating units to control and intervention groups. Sample sizes reflect a calculation of the statistical power needed for any programme effect identified by the evaluations to be greater than chance. Robust measurement was also required. These elements are typical of a good RCT. The focus on replicating findings from other trials offers a different angle, however. Specifically, the data collection was restricted to the factors in the logic model underpinning the evidence-based programme, including the risks targeted, the fidelity of implementation of core elements of the intervention, and the outcomes sought. Other hypothesised moderators and other contextual information are

excluded. The net result is a high-quality evaluation with less data and therefore less cost.

### 3. Evaluation

Family Nurse Partnership (FNP) was introduced as part of a national evaluation. The largest RCT of FNP is being undertaken by the University of Cardiff and will report in 2013. Evaluations of Incredible Years, PATHS and Triple-P by RCT were undertaken by the Social Research Unit in collaboration with the Peninsula Medical School, which provided statistical advice, and the Centre for Health Economics and Medicines Evaluation at Bangor University, which examined the cost-effectiveness of Incredible Years, Triple-P and PATHS. Process and qualitative data were collected alongside the trials. This article presents a summary of outcome results from the three trials, with brief reference to fidelity of implementation where it may explain results (fidelity is covered in more depth in papers on each of the trials, to follow).

The Incredible Years evaluation was a parallel RCT with pre-post test design, which involved the parents of 161 children aged three and four, identified through referral from other agencies, self-referral, and screening of children served by children’s centres. In order to be eligible for the programme children needed to be at risk of a social-emotional or behavioural disorder, which meant reaching the “high need” threshold (17 or above out of 40) of the “total difficulties” score of the parent-completed Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997). There were 101 males and 60 females, with a mean age of 44 months (SD = 6) at baseline. The sample comprised a high proportion of low-income families: 50 percent of families relied on benefits as their main source of income.

The 161 children who reached the threshold were randomised to the study on a 2:1 ratio – 110 to intervention, 51 to control. Baseline data (Wave 1) were collected on all 161 children before randomisation. Follow-up (Wave 2) took place six months after baseline.<sup>1</sup> It included 147 children,

<sup>1</sup> It was possible to conduct a long-term follow-up at one-year post-baseline (Wave 3) as well but this was for the intervention group only, not the control group. Results are reported elsewhere.



leaving 14 missing cases (10 intervention and four control): seven formal withdrawals from the study and seven who could not be contacted. As with all three trials, values were imputed for missing data, based on baseline scores. The SDQ and the Eyberg Child Behaviour Inventory (ECBI) (Eyberg and Ross 1978), both completed by parents, were used to measure child outcomes. Parenting behaviour was measured using the Arnold and O’Leary Parenting Scale (APS) (Arnold et al. 1993). As Table 1 shows, there were differences between intervention and control groups at baseline. The impact of Incredible Years is therefore calculated using an estimated mean difference between intervention and control groups. ANCOVA tests controlled for children’s start scores on the respective measure, the age and sex of the child, and the centre from which families were recruited.

The PATHS evaluation was a cluster-randomised design involving 56 schools, 29 of which were allocated to implement the PATHS curriculum and 27 of which were allocated to a waiting list. The Bangor Trials Unit (NORTH) generated the randomisation sequence and the sample was stratified by percentage of free school meals and size of school. As a universal intervention PATHS was offered to all children in reception and year one classes (aged four–six years). This cohort received the programme for two years. There were 5,397 pupils in the schools at baseline, attending 196 classes (102 intervention and 94 control). Data were collected at three points: baseline (Wave 1) in September 2009; first follow-up (Wave 2) in June 2010 (after one year of implementation); and second follow-up (Wave 3) in June 2011 (on completion of the intervention). Data were collected on 183 classes ( $n = 5,074$ ) at Wave 1, on 176 classes ( $n = 4,998$ ) at Wave 2, and on 178 classes ( $n = 4,994$ ) at Wave 3. There were 4,006 complete cases with data at all three waves.

Outcomes were assessed using the SDQ teacher report and the PATHS Teacher Rating Survey (PTRS), a composite measure of seven scales (e.g. Child Behaviour Questionnaire). These two measures provide a picture of

children’s behavioural and emotional difficulties, social competence, and emotional regulation. To identify the unique contribution of PATHS to the outcomes measured, hierarchical linear models were run to take account of class/school clustering in outcomes and also to control for available covariates at all levels. Data were first analysed using only complete cases, ignoring missing data. A second set of analyses used multiple imputation methods to account for missing data. However, this did not fundamentally change the results.

The evaluation of Triple-P was a parallel randomised controlled trial, with pre-post test design. It involved 146 children aged four–nine years whose symptoms indicated a potential social-emotional or behavioural disorder, determined using the “high need” threshold on the SDQ “total difficulties” score (17 or above out of 40). The sample comprised 105 boys and 41 girls. The mean age was 82 months ( $SD = 21$ ). The sample also comprised a high proportion of low-income families: 62 percent of children were entitled to free school meals compared to 33 percent for Birmingham as a whole.

The parent(s) of half (73) of these children were randomly assigned to attend Triple-P parenting groups, with the remaining half placed on a waiting list and receiving services as usual. Researchers performed the randomisation for each eligible child using an online programme, designed by NORTH. Children were randomised on a 1:1 ratio, using a dynamic allocation method, stratified by age and sex. Baseline (Wave 1) data was collected on all children. Follow-up (Wave 2) occurred six months after baseline and included 137 children.<sup>2</sup> The programme was delivered to intervention group parents at some point during those six months. The missing nine cases (three control, six intervention) were made up of two formal withdrawals from the study and seven that could not be contacted. The primary outcome instruments were the SDQ and ECBI. Parenting behaviour was measured using the Arnold and O’Leary Parenting Scale (APS). Estimated mean differences

<sup>2</sup> It was possible to conduct a long-term follow-up at one-year post-baseline (Wave 3) as well but this was for the intervention group only, not the control group. Results are reported elsewhere.

were used to calculate the impact of Triple-P. ANCOVA tests controlled for children's start scores on respective measures, the age and sex of the child, and the area from which families were recruited.

#### 4. Results

The three evaluations revealed both success and failures for the Birmingham Brighter Futures strategy.

#### 4.1. Incredible Years

The results summarised in Table 1 indicate significant benefits from the Incredible Years programme offered as part of children's centre provision (universal early years). There are reductions in reported negative parenting behaviours among the parents attending Incredible Years groups compared to controls. There are strong reductions in child behaviour problems and improvements in relationships for children receiving the intervention.

**Table 1: Child outcomes for Incredible Years control and intervention groups**

Mean (SD) raw scores						
Child measure (cut-off)	Control (n=51)		Intervention (n=110)		Estimated mean difference (95% CI) <sup>1</sup>	Effect Size (d)
	Baseline	6 mth	Baseline	6 mth		
SDQ conduct <sup>2</sup> problems (4)	6.53 (2.1)	4.43 (2.7)	6.29 (2.0)	3.62 (2.1)	0.78* (0.05 to 1.51)	0.39
SDQ emotion problems (5)	4.85 (2.5)	3.61 (2.6)	4.79 (2.4)	3.30 (2.3)	0.36 (-0.36 to 1.07)	
SDQ hyperactivity (7)	7.66 (1.9)	6.18 (2.4)	7.67 (1.9)	5.83 (2.5)	0.40 (-0.36 to 1.17)	
SDQ peer problems (4)	4.47 (1.9)	3.39 (2.1)	4.23 (1.8)	2.69 (1.8)	0.71* (0.85 to 1.34)	0.39
SDQ pro-social behaviour (<4) <sup>3</sup>	5.19 (2.2)	6.35 (2.2)	5.72 (2.1)	6.77 (2.1)	-0.22 (-0.84 to 0.40)	
SDQ total difficulties (17)	23.50 (4.5)	17.60 (7.3)	22.98 (4.4)	15.44 (6.0)	2.23* (0.13 to 4.34)	0.50
SDQ impact (2)	0.92 (1.4)	0.58 (1.2)	0.59 (1.1)	0.14 (0.5)	0.37** (0.10 to 0.63)	0.31
ECBI-I (127) <sup>4</sup>	143.86 (38.5)	134.35 (42.3)	142.70 (35.7)	123.10 (34.8)	13.48* (2.31 to 22.64)	0.37
ECBI-P (11) <sup>5</sup>	17.31 (9.3)	14.33 (9.8)	16.71 (8.8)	11.24 (9.0)	2.62 (-0.07 to 5.32)	
APS total <sup>6</sup>	3.58 (0.8)	3.32 (0.8)	3.49 (0.6)	3.01 (0.8)	0.29** (0.08 to 0.50)	0.43
APS laxness	3.79 (1.3)	3.43 (1.2)	3.58 (1.2)	3.04 (1.1)	0.30 (-0.01 to 0.61)	
APS verbosity	4.15 (0.9)	4.01 (1.0)	4.26 (0.9)	3.68 (1.0)	0.42** (0.12 to 0.72)	0.47
APS over-react	2.90 (1.0)	2.71 (1.1)	2.78 (0.8)	2.36 (0.8)	0.31* (0.06 to 0.57)	0.36

\* significant at  $p < .05$

\*\* significant at  $p < .01$

1 Difference in mean follow-up scores between intervention and waiting list control conditions, measured by analysis of covariance adjusted for baseline score, age of child, sex and area.

2 SDQ Strengths and Difficulties Questionnaire (on all scales higher scores equals greater problems, except for prosocial behaviour).

3 SDQ prosocial is measured positively: the higher the score, the better the behaviour of the children.

4 ECBI-I Eyberg Child Behaviour Inventory - Intensity Scale (higher scores equate to greater problems).

5 ECBI-P Eyberg Child Behaviour Inventory - Problem Scale (higher scores equate to greater problems).

6 APS Arnold and O'Leary Parenting Scale (and three sub-scales) (higher scores equate to greater problems).

The data suggest an effect size on SDQ conduct problems of 0.39, consistent with the 0.33 reported by Hutchings and colleagues (2007a) for their similar trial in Wales. In common with other evaluations of Incredible Years, benefits increased with dose. Training and supervision arrangements for Incredible Years facilitators, which boost fidelity of programme delivery, were also implicated in better outcomes. The effect size of 0.37 for the ECBI-I is smaller than in other UK community-based trials – 0.89 found by Hutchings et al. (2007) and 0.55 by Gardner et al. (2006) – but is nevertheless respectable, particularly given that implementation in Brighter Futures was arguably more “real world” than in these two earlier studies.

#### 4.2. PATHS

The results for PATHS were more mixed. Table 2 presents the results from the hierarchical linear model analysis, examining the difference in outcomes between the PATHS group and control group, accounting for class- and school-level clustering in outcomes and available covariates at all levels. As indicated, at first follow-up there were modest improvements in pupils’ emotional health and behavioural development in the PATHS schools compared to those in control schools. However, at the two-year follow-up, these gains had all been lost.

**Table 2: Hierarchical linear model results for the PATHS programme with N and between-group difference mean (95% CI)<sup>1</sup>**

	12-month follow-up adjusted for baseline	24-month follow-up adjusted for baseline
SDQ total difficulties	N=4255 -0.42 (-1.11 to 0.28)	N=3934 0.19 (-0.64 to 1.03)
SDQ impact	N=4123 -0.34 (-0.11 to 0.05)	N=4000 0.04 (-0.05 to 0.12)
SDQ conduct	N=4265 -0.15 (-0.31 to 0.01)	N=3953 0.16 (-0.04 to 0.35)
SDQ emotion	N=4265 -0.12 (-0.33 to 0.10)	N=3953 0.06 (-0.18 to 0.30)
SDQ pro-social	N=4265 0.18 (-0.16 to 0.52)	N=3953 0.16 (-0.27 to 0.59)
PTRS emotional regulation	N=4203 0.11 (-0.04 to 0.27)	N=4019 -0.18 (-0.35 to 0.00)
PTRS pro-social behaviour	N=4203 0.16 (-0.01 to 0.32)	N=4019 -0.06 (-0.25 to 0.13)
PTRS social competence	N=4203 0.14 (0.01 to 0.29)*	N=4019 -0.11 (-0.29 to 0.01)
PTRS aggressive	N=4203 -0.13 (-0.23 to -0.04)*	N=4019 0.01 (-0.09 to 0.13)
PTRS internalising	N=4203 -0.16 (-0.27 to -0.04)	N=4009 0.01 (-0.12 to 0.22)
PTRS hyperactive behaviour	N=4226 -0.07 (-0.13 to -0.001)*	N=4040 0.03 (-0.04 to 0.11)
PTRS peer problems	N=4217 -0.12 (-0.22 to -0.02)*	N=4003 0.08 (-0.04 to 0.21)
PTRS relational aggression	N=4217 -0.08 (-0.20 to 0.01)	N=3998 0.05 (-0.07 to 0.16)
PTRS learning behaviours	N=4180 0.05 (0.003 to 0.10)*	N=3974 -0.01 (-0.07 to 0.05)

\* indicates a significant difference between PATHS and control ( $p < .05$ )

<sup>1</sup> A negative mean score indicates the follow-up score in the intervention group is lower than in the control group for that construct. Measures reporting on positive behaviours, where higher scores equate to better outcomes, are shaded grey on the tables. All other measures are reported negatively, higher score equates to greater problems. The ICC values at the level of the classroom on SDQ constructs ranged between 0.07 and 0.28.



There were some sub-group differences. The benefits of PATHS increased with age, although not significantly. There was a significant impact at two years on students who tested as depressed and/or anxious at baseline. White students benefited more than other ethnic groups, though not significantly so. Poverty did not emerge as a moderator of results.

#### 4.3. Standard Level-4 Triple-P

As Table 3 illustrates, the results for this programme are not promising. Children of parents attending Triple-P

sessions improved their behaviour and were happier six months after the course concluded, but at roughly the same rate as children in the control group receiving services as normal. These results are not consistent with most other Triple-P trials around the world. However, as far as we are aware, only four randomised trials (including this one) have been undertaken independent of the programme originator (see also Gallart and Matthey 2005; Hahlweg et al. 2010; Malti, Ribeaud, and Eisner 2011). When these four studies are viewed together, the evidence of impact on child development is equivocal.

**Table 3: Child outcomes for Triple-P control and intervention groups**

Mean (SD) raw scores						
Child measure (cut-off)	Control (n=73)		Intervention (n=73)		Estimated mean difference (95% CI)	p value
	Baseline	6 mth	Baseline	6 mth		
SDQ conduct problems (4)	5.28 (2.0)	4.13 (2.2)	5.42 (2.3)	4.33 (2.2)	-0.15 (-0.79 to 0.50)	0.65
SDQ emotion problems (5)	5.19 (2.9)	4.25 (2.8)	5.40 (2.4)	4.20 (2.8)	0.84 (-0.70 to 0.86)	0.83
SDQ hyperactivity (7)	8.13 (1.8)	7.08 (2.5)	7.66 (1.9)	6.52 (2.4)	0.21 (-0.48 to 0.90)	0.54
SDQ peer problems (4)	4.29 (2.1)	3.63 (2.1)	4.75 (1.8)	3.97 (2.3)	-0.12 (-0.77 to 0.53)	0.72
SDQ pro-social behaviour (<4)	5.70 (2.1)	6.56 (2.0)	5.98 (2.4)	6.81 (2.2)	-0.10 (-0.63 to 0.43)	0.71
SDQ total difficulties (17)	22.89 (4.3)	19.09 (7.0)	23.23 (4.4)	19.02 (7.6)	0.22 (-1.78 to 2.21)	0.83
SDQ impact (2)	3.99 (2.5)	2.73 (2.8)	4.67 (2.9)	3.01 (3.4)	-0.09 (-1.01 to 0.84)	0.85
ECBI-I (127)	156.07 (39.8)	141.51 (43.2)	155.00 (38.8)	143.64 (45.3)	-4.39 (-14.64 to 5.86)	0.40
ECBI-P (11)	20.47 (8.5)	15.89 (9.8)	19.96 (8.3)	15.96 (9.5)	-1.97 (-4.28 to 0.35)	0.10
APS total	3.62 (0.7)	3.34 (0.7)	3.62 (0.7)	3.29 (0.9)	0.05 (-0.16 to 0.26)	0.65
APS laxness	3.53 (1.2)	3.13 (1.1)	3.52 (1.3)	3.15 (1.3)	-0.45 (-0.37 to 0.28)	0.79
APS verbosity	4.40 (0.7)	4.23 (1.0)	4.41 (0.9)	4.19 (1.0)	0.04 (-0.26 to 0.33)	0.81
APS over-reactivity	3.17 (1.0)	2.92 (0.9)	3.12 (0.9)	2.77 (1.1)	0.15 (-0.14 to 0.43)	0.32

\* significant at  $p < .05$

\*\* significant at  $p < .01$

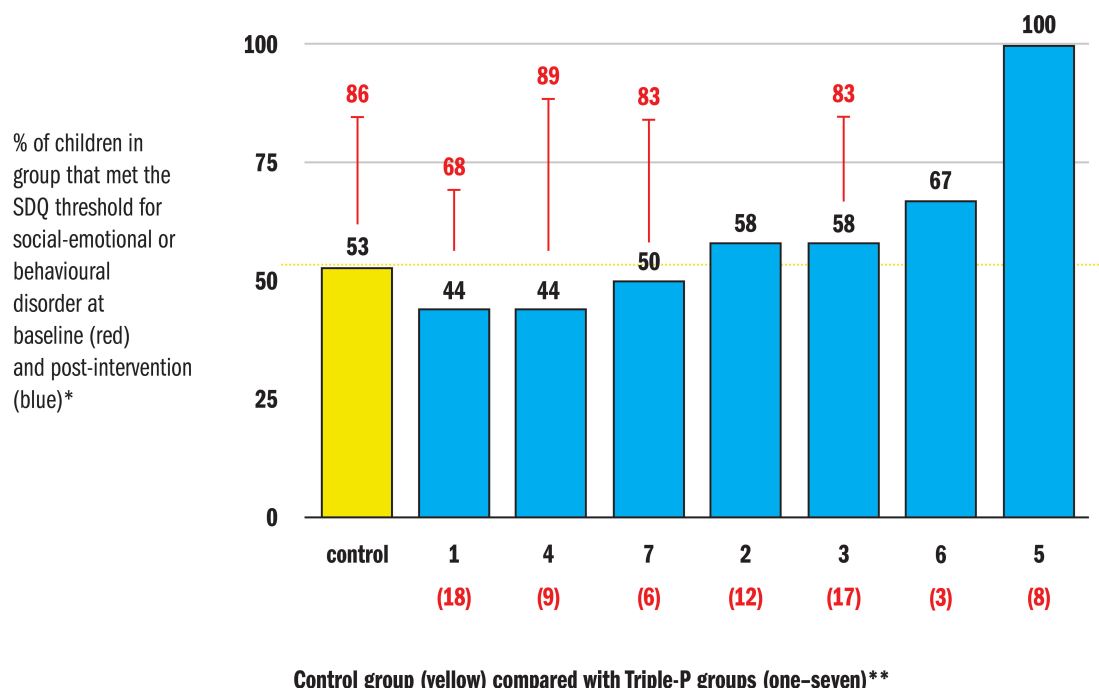
Poor fidelity of implementation may explain the failure to detect an impact on child development in the Birmingham Triple P trial (this is examined in more depth elsewhere). Brighter Futures inherited existing Triple-P practitioners who received “top-up” training from programme originator Matt Sanders. There was considerable variability in

the quality of provision and results achieved by these Triple-P practitioners. Not all parents received their full dose of Triple-P, with an average attendance across the groups of 40 percent. As Table 4 illustrates, one practitioner (group four) managed to reduce conduct disorders in the children of parents she or he was working with by 45 per

cent, while others failed to achieve any change. Emotional disorders for children whose parents participated in groups

led by three practitioners got worse, not better, than those in the control group receiving services as usual.

**Table 4: Group-level outcomes for Triple-P**



\* Where a number only appears in blue (group two, five and six) the baseline and post-intervention % are identical

\*\* Numbers in parentheses refer to number of parents

## 5. Discussion

The Brighter Futures work in Birmingham is a good example of how it is possible to create a “pull” for science to be used in social policy. Most attempts to link science and policy depend on “push”, injecting evidence into reluctant public systems. The *Common Language* work led Birmingham to demand and embed strong epidemiological data on the well-being of local children, good international evidence on “what works”, and reliable indications concerning the costs and benefits of competing investment options (Axford and Morpeth 2012). Perhaps most impressively, Birmingham committed to four experimental trials of new interventions. The local authority needed to know if these innovations worked. It would be surprising to find four

local authority-sponsored experimental evaluations in all of the other 149 English local authorities combined.

The resulting strategy had many positive features. It was proactive not reactive. It sought to forestall future problems rather than fire-fighting existing problems. It focused more on prevention and early intervention than on treatment, and it took a developmental approach to children. It tried to prevent children born into high-risk households falling behind. It sought to shift the distribution for all primary school children’s social and emotional regulation (see Rose 1998).

The strategy was unique at the time of its preparation in giving emphasis to evidence-based programmes proven by

the highest standards to work. And Brighter Futures introduced a new standard of financial accountability, seeking not only to invest in children's development but also to internalise the collection of high quality programme cost data at programme, school, group, or child level, as appropriate, with a view to producing estimates of short-term cost effectiveness and long-term return on investment.

The strategy had unanticipated benefits. The epidemiology had suggested that 15 percent of children aged three–four in the city would fall into the “high need” category of the SDQ “total difficulties” score. However, children's centres struggled to find these children in their catchment areas. Of an estimated pool of 437 children meeting that threshold in the relevant catchment areas, centres (collectively) needed to recruit 144 for the RCT but were only able to find 89 in the original recruitment period. This led to a concerted and ultimately successful effort to boost recruitment, including outreach and financial incentives, and taught important lessons about how children's centres can reach more disadvantaged families (Axford et al. 2012). This was valuable in the context of wider policy discussions about how children's centres need to refocus their activity.

In these and other respects Brighter Futures was radical and forward looking, and Birmingham should be commended for going first where many other local authorities are now hesitantly following. The leadership of the Chief Executive, elected politicians and the Strategic Director of Children's Services was a fundamental component in the success of Brighter Futures.

But there were also many limitations in the work. In the space available just two will be discussed. First, Brighter Futures started as and never quite progressed beyond the status of a “project” or “pilot”. It was a marginal not mainstream activity. It was a big project in a big pool. Over £40 million was set aside for Brighter Futures, but Birmingham was spending over £1.3 billion annually on its 260,000 children.

The marginal quality had significant effects when, almost inevitably in a local authority of its size, a preventable child death occurred (Radford 2010). In the political and media focus on this case, Brighter Futures was seen as a hindrance

– in the sense of significant resources being spent on prevention and early intervention rather than child protection in the traditional sense – when in other circumstances it might have been viewed as fundamental to righting the problem. Family Nurse Partnership (FNP), for example, is the most effective intervention known for preventing child maltreatment (e.g. MacMillan et al. 2009)

Second, the evidence-based programmes introduced by Brighter Futures had different impacts. Incredible Years was a success, as was PATHS initially (after one year of implementation). But after two years of implementation, PATHS only had an appreciable impact on children with emotional disorders. The Standard Level-4 Triple-P parenting programme not only had zero benefits overall but when poorly attended it generated potentially iatrogenic effects. The longer-term effects of these programmes – for example, six or twelve months after the programme ended – were not studied.

There is not space here to explore this variation fully. Although evidence-based programmes work, the size of effect is frequently small and can be diminished to nothing. In the case of Triple-P, fidelity of implementation was a problem. It is not sufficient simply to introduce an evidence-based programme; it has to be put into practice with great care and effort (Fixsen et al. 2005).

It is possible the local context may play a part in reducing impact of evidence-based programmes. PATHS, for example, has been proven to work in poorly funded schools in the United States serving high-risk communities (e.g. CPPRG 2010). European schools are invariably better funded, and so the existing provision – “services as usual” – against which programmes like PATHS are compared, might be stronger than in the US. In England, for instance, a project supported by national government, Social and Emotional Aspects of Learning (SEAL), had been addressing primary school pupils' social and emotional regulation for over five years. When Ross and colleagues (2012) evaluated PATHS in Northern Ireland they found only marginal benefits and an evaluation of the programme in Switzerland has produced similar results to those reported here (Malti et al. 2011).

It is possible that the different populations that the programmes served is part of the explanation for Incredible Years' success compared with Triple-P and PATHS. In Birmingham, Incredible Years was targeted at three–four years olds, while Triple-P was used for children aged four–nine years. Older children may have more entrenched difficulties. In addition, PATHS was delivered as a universal intervention to all classroom children; it is possible that greater effects may have been found with a targeted population.

It is known that the involvement of the programme originator in evaluations often skews the results in a positive direction (Eisner 2009). The four Birmingham trials were undertaken independent of the programme originators.

Another hypothesis is that some evidence-based programmes are more transportable than others. The consistent and positive findings for Incredible Years across contexts are striking. It works in the United States, Ireland, Wales, London, Birmingham and many other contexts worldwide. If transportability is a problem, what does Incredible Years have that other evidence-based programmes do not? This said, other studies show that Triple-P transports well in terms of impact, albeit – as indicated earlier in this article – with the involvement of the programme originator in the evaluation.

At this stage we can only speculate on reasons for the variability of results. What can be said for certain is that evidence-based programmes are a stepping-stone to future

improvement to children's health and development. They are not a panacea.

Although we were successful in applying a standard cost-effectiveness approach to each of the three trials, they could not be compared directly against one another due to differing ages and outcome measures. We calculated the cost per child for each programme and, in the case of Incredible Years, found an incremental cost-effectiveness ratio in line with previous economic evidence for this programme (Edwards et al. 2007). However, it has yet to be established whether there was any longer term return on investment from the Brighter Futures initiative in terms of benefits to the judicial system, education system or health care system.

In conclusion, Brighter Futures was a brave experiment. It demanded better science in policy formulation and involved testing evidence-based programmes in real world settings. Other studies had done this previously, including in relation to Incredible Years (Gardner et al. 2006; Hutchings et al. 2007a/b), but Brighter Futures went further along the efficacy-effectiveness spectrum by testing implementation city-wide in a range of regular services without a prior tradition of implementing these evidence-based programmes. It will leave an important legacy for the city, its children and other large-scale systems in Europe. Much can be learned from the Birmingham experience by those engaged in continuing attempts to improve children's health and development.

## References

- Aos, Steve, Roxanne Lieb, Jim Mayfield, Marna Miller, and Annie Pennucci. 2004. *Benefits and Costs of Prevention and Early Intervention Programs for Youth*. Olympia: Washington State Institute for Public Policy.
- Arnold, David S., Susan G. O'Leary, Lisa S. Wolff, and Maureen M. Acker. 1993. The Parenting Scale: A Measure of Dysfunctional Parenting in Discipline Situations. *Psychological Assessment* 5: 137–44.
- Axford, Nick, Louise Morpeth, Michael Little, and Vashti Berry. 2008. Linking Prevention Science and Community Engagement: A Case Study of the Ireland Disadvantaged Children and Youth Programme. *Journal of Children's Services* 3 (2): 40–54.
- Axford, Nick, and Tim Hobbs. 2011. Getting the Measure of Child Health and Development Outcomes (1): A Method for Use in Children's Services Settings. *Child Indicators Research* 4 (1): 81–100.
- Axford, Nick, David Jodrell, and Tim Hobbs. 2012. Making Child Well-being Data Work Hard: Getting from Data to Policy and Practice. *Child Indicators Research*. doi:10.1007/s12187-012-9163-5
- Axford, Nick, Minna Lehtonen, Kate Tobin, Dwan Kaoukji, and Vashti Berry. 2012. Engaging Parents in Parenting Programs: Lessons from Research and Practice. *Children and Youth Services Review* 34 (10): 2061–71.
- Axford, Nick, and Louise Morpeth. 2012. The Common Language Prevention Operating System: From Strategy Development to Implementation of Evidence-based Practice. In *Handbook of Implementation Science for Educational Psychology*, ed. Barbara Kelly and Daniel F. Perkins, 443–58. Cambridge: Cambridge University Press.
- BCC (Birmingham City Council). 2007. *A Brighter Future for Children and Young People: The Birmingham Strategy*. [http://ebriefing.bgfl.org/bcc\\_ebrief/content/resources/resource.cfm?id=4314&key=&zz=20111021114309327&zs=n](http://ebriefing.bgfl.org/bcc_ebrief/content/resources/resource.cfm?id=4314&key=&zz=20111021114309327&zs=n)
- Bodenman, Guy, Annette Cina, Thomas Ledenmann, and Matthew R. Sanders. 2008. The Efficacy of Positive Parenting Program (Triple P) in Improving Parenting and Child Behavior: A Comparison with Two Other Treatment Conditions. *Behaviour Research and Therapy* 46 (4): 411–42.
- Collishaw, Stephan, Barbara Maughan, Robert Goodman, and Andrew Pickles. 2004. Time Trends in Adolescent Mental Health. *Journal of Child Psychology and Psychiatry* 45 (8): 1350–62.

- Collishaw, Stephan, Barbara Maughan, Lucy Natarajan, and Andrew Pickles. 2010. Trends in Adolescent Emotional Problems in England: A Comparison of Two National Cohorts Twenty Years Apart. *Journal of Child Psychology and Psychiatry* 51 (8): 885–94.
- CPPRG (Conduct Problems Prevention Research Group) 2010. The Effects of a Multiyear Universal Social-emotional Learning Program: The Role of Student and School Characteristics. *Journal of Consulting and Clinical Psychology* 78 (2): 156–68.
- Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger. 2011. The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis Of School-Based Universal Interventions. *Child Development* 82 (1): 405–32.
- Edwards, Rhiannon Tudor, Alan Ó Céilleachair, Tracey Bywater, Dyfrig A. Hughes, and Judy Hutchings. (2007) A Parenting Programme for Children at Risk of Developing Conduct Disorder: A Cost-Effectiveness Analysis. *British Medical Journal* 334: 682.
- Eisenstadt, Naomi. 2011. *Providing a Sure Start: How Government Discovered Early Childhood*. Bristol: Policy Press.
- Eisner, Manuel. 2009. No Effects in Independent Prevention Trials: Can We Reject the Cynics' Hypothesis? *Journal of Experimental Criminology* 5:163–83.
- Eyberg, Sheila M., and Arthur W. Ross. 1978. Assessment of Child Behavior Problems: The Validation of a New Inventory. *Journal of Clinical Child Psychology* 7:113–16.
- Fixsen, Dean L., Sandra F. Naoom, Karen A. Blase, Robert M. Friedman, and Frances Wallace. 2005. *Implementation Research: A Synthesis of the Literature*. Tampa: USE.
- Gardner, Frances, Jennifer Burton, and Ivana Klimes. 2006. Randomised Controlled Trial of a Parenting Intervention in the Voluntary Sector for Reducing Child Conduct Problems: Outcomes and Mechanisms of Change. *Journal of Child Psychology and Psychiatry* 47:1123–32.
- Gallart, Sonia C., and Stephen Matthey. 2005. The Effectiveness of Group Triple P and the Impact of the Four Telephone Contacts. *Behaviour Change* 22 (2): 71–80.
- Goodman, Robert. 1997. The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry* 38 (5): 581–86.
- Greenberg, Mark T., and Carol Kusché. 2002. *Promoting Alternative Thinking Strategies (PATHS)*, Blueprints for Violence Prevention 10. Boulder: Center for the Study and Prevention of Violence, University of Colorado.
- Hahlweg, Kurt, Nina Heinrichs, Annett Kuschel, Heike Bertram, and Sebastian Naumann. (2010). Long-term Outcome of a Randomized Controlled Universal Prevention Trial through a Positive Parenting Program: Is it Worth the Effort? *Child and Adolescent Psychiatry and Mental Health* 4:14–27.
- Hobbs, Tim, Nick Axford, and David Jodrell. 2011. Getting the Measure of Child Health and Development Outcomes (2): The Picture for a Local Authority in England. *Child Indicators Research* 4 (1): 81–100.
- Hutchings, Judy, Tracey Bywater, Dave Daley, Frances Gardner, Chris Whitaker, Karen Jones, Catrin Eames, and Rhiannon T. Edwards. 2007. Parenting Intervention in Sure Start Services for Children at Risk of Developing Conduct Disorder: Pragmatic Randomized Controlled Trial. *British Medical Journal* 334 (7595): 1–7.
- Hutchings, Judy, Tracey Bywater, and Dave Daley. 2007. Early Prevention of Conduct Disorder: How and Why did the North and Mid Wales Sure Start Study Work? *Journal of Children's Services* 2 (2): 4–14.
- Little, Michael, and Ali Abunimah. 2007. Improving Outcomes for Children in the Island of Ireland: The Role of Philanthropic Investment. *Journal of Children's Services* 2 (2): 60–67.
- MacMillan, Harriet L., C. Nadine Wathen, Jane Barlow, David M. Fergusson, John M. Leventhal, and Heather N. Taussig. 2009. Interventions to Prevent Child Maltreatment and Associated Impairment. *Lancet* 373 (9659): 250–66.
- Malti, Tina, Denis Ribeaud, and Manuel P. Eisner. 2011. The Effectiveness of Two Universal Preventive Interventions in Reducing Children's Externalizing Behavior: A Cluster Randomized Controlled Trial. *Journal of Clinical Child & Adolescent Psychology* 40 (5): 677–92.
- Maughan, Barbara, Stephan Collishaw, Howard Meltzer, and Robert Goodman. 2008. Recent Trends in UK Child and Adolescent Mental Health. *Social Psychiatry and Psychiatric Epidemiology* 43: 305–10.
- Olds, David L., Peggy G. Hill, Sharon F. Mihalic, and Ruth A. O'Brien. 1998, updated 2006. *Prenatal and Infancy Home Visitation by Nurses: Blueprints for Violence Prevention*. Boulder, CO: Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado.
- Radford, John. 2010. *Serious Case Review Under Chapter VIII "Working Together to Safeguard Children" In respect of the Death of a Child, Case Number 14*. Birmingham: Birmingham Safeguarding Children Board.
- Rose, Geoffrey. 1998. *Rose's Strategy of Preventive Medicine*. Oxford: Oxford University Press.
- Ross, Steven M., Mary K. Sheard, Alan Cheung, Louise Elliott, and Robert Slavin. 2012. Promoting Primary Pupils' Social-emotional Learning and Pro-social Behaviour: Longitudinal Evaluation of the Together4All Programme in Northern Ireland. *Effective Education* 3 (2): 61–81.
- UNICEF. 2007. *Child Poverty in Perspective: An Overview of Child Wellbeing in Rich Countries*. Florence: UNICEF Innocenti Research Centre.
- Webster-Stratton, Carolyn. 1994. Advancing Videotape Parent Training: A Comparison Study. *Journal of Consulting and Clinical Psychology* 66: 715–30.

**Michael Little**  
 mlittle@dartington.org.uk

**Vashti Berry**  
 vberry@dartington.org.uk

**Louise Morpeth**  
 lmorpeth@dartington.org.uk

**Sarah Blower**  
 sblower@dartington.org.uk

**Nick Axford**  
 naxford@dartington.org.uk

**Rod Taylor**  
 rod.taylor@pms.ac.uk

**Tracey Bywater**  
 tracey.bywater@york.ac.uk

**Minna Lehtonen**  
 mlehtonen@dartington.org.uk

**Kate Tobin**  
 ktobin@dartington.org.uk